Sentimental Analysis Using Web Media

Jeevan Pendam[1], Aman Mishra[2], Satyam Mishra[3], S. P. Pawar[4]

[1,2,3]*Student, Department of Information Technology, A. C. Patil College of Engineering, Kharghar, India*
[4]*Professor, Department of Information Technology, A. C. Patil College of Engineering, Kharghar, India*

Abstract: **The quality of a product is an important factor which is used to determine rating of a product in recent days. Hidden sentiment in review and rate of product is detected by the product rating system. Products which are displayed on e-commerce website in which users or customer give the review and according to that it gets analyzed and in order to achieve high functionality, system uses sentiment analysis methodology. Accurate rating which are generate by user's review is because of a system which completely destroy the trouble of giving rating and write review. The output of proposed system is e-commerce website where users are giving review on a particular product and later these reviews are passed to data cleaning and preprocessing and based on that rating is generated. The rating of a product which are generated is out of 5 and the average rating of product is consist of all individual rating.**

Keywords: **Sentiment Analysis, Product Reviews, Rating, Users, E-commerce.**

1. Introduction

Now-a-days internet is the most valuable source of learning getting reviews of product or discovering new ideas. Everyday millions of reviews are generated on specific product because of their huge number and size it is difficult to handle. Natural Language Processing, Text Analysis and Text Preprocessing are the major research field in the current time. Different technique and tools are used to analyze the sentiment.

A. Existing System

Sentiment classification technique divide into two category machine learning approach and lexicon approach [2]. The lexicon based approach depends upon sentiment lexicon and set of commonly known and a precompiled sentiment word. In lexicon based approach there is a predefined dictionary of vocabulary with its score and it could be the polarity. In first phase which is a development phase user need to be made a predefined dictionary with a score of all possible adjective and adverb which have the possibilities to be used maximum the words added increases the accurate result. In second phase the several rules were applied to obtain the results.

For eg: I love the product. = there will be score of word love which can be 3 the final score will become 3.

B. Problem Statement

The ever-increasing use of e-commerce [1] has produced vast consumer generated contents such as reviews. It is becoming imperative for organizations as well as consumers to collect these data for product analysis or rating according to user needs.

C. Proposed system

As per the paper which are present in literature review, a bit of idea is generated is to determine the rating of products in an ecommerce website by analyzing the user reviews. Proposed system is capable to determining new reviews taking reference to previously trained reviews. In this system by using real time comment and after analysis of that it will display the result in rating. A Traditional Sentiment Analysis system has some training data, which is then fed to the classifier and the result is obtained.

- The Inputs: Building an E-commerce website and taking the user comment as per the purchased product. Then this comment will get stored with rating in" MySQL" database and it will retrieve on website with rating.
- Data pre-processing and Cleaning: At first, data is splitted into train sentence and labels. Afterwards data will convert into lowercase. The process through which it can be done that are mentioned below:
  - o Splitting the sentence
  - o Removing special characters
  - o Obtain the sentiment.
- Splitting the sentence: In this, the word is getting splitted from the sentence so that it will become easier to train and to use/take it as input through which, result is obtained.
- Removing Special Characters: In this section, special character such as @,#,$,etc. are removed from the sentence and will use for preprocessing of data.
- Obtain the Sentiment: From above two section it will be clear that how to preprocess the data and how to feed in the reviews in the LSTM [6] network.
- Text to sequence: It will transform each text into sequence of integer.
- Pad sequence: In this it will transform list of number sample sequence into a matrix of shape.

2. Literature analysis

- Sentimental Analysis on online product review [2017]: In this paper, Raheesa Safrin, E.A. Vimal used feature vector classification for analysis of sentiment. The data used in this study is online product review collected from sample website that we have created word such as an adjective, adverb are able to convey opposite sentiment with the

International Journal of Research in Engineering, Science and Management
Volume-3, Issue-2, February-2020
www.ijresm.com | ISSN (Online): 2581-5792

464

help of negative prefix, negative phrase identification is used.

- Sentimental Analysis of Product review [2014] : In this paper, Aarti Patil used advanced Naïve Bayesian Algorithm technique to find whether the reviews are positive or negative for analysis of sentiment. The sentiment analysis is used to extract, aggregate and analyse the opinion on product from discussion forum.
- Comparative Analysis of Sentimental Orientation using SVM and Naïve Bayes Techniques (2016): In this paper, Shweta Rana used Naïve Bayes classifier in system. Sentiment classified as positive or negative sentiment using film user review. Algorithm like Naïve Bayes, Linear SVM and Synthetic words is used.
- Extracting a sentiment from Review: A Lexicon Based Approach[2017] : In this paper, Sujuta Sonavane used Lexicon Based Approach for extracting sentiment such as positive, negative or neutral. Extracting the useful content from the opinion sources becomes a challenging task. To extract sentiment from review SentiWordNet is used to assign the polarity of sentiment. The classification of review document is predicted by sentimental score.

3. Description of Modules

In this section, it will tell about the working of a system in which it tells how it is work such as how to take an input and preprocess that input to obtain the sentiment of review.

A. Working

The project work is in the three phase:
- Creating website and retrieving the feedback
- Data Cleaning and Pre-processing
- Obtain the sentiment and analyze the data

Creating website and retrieving the feedback: Building an E-commerce website and taking the user comment as input. Sentiment classification is used to verify or analyze the comments given by the user to extract the opinion. The Front End is developed using languages like HTML and the design to the site is given using CSS. Python is used in the server side scripting to perform the connection to database get message from user and feed to LSTM model.

First step is to train the model using database for eg. IMDB in which input data will be train the some of the sentence based on their polarity and then data will stored on this model with extension using h5? for high and valuable data get stored, because it is an long process and take more than a day for training sentence to save time, the model is trained and save it for using again and again.

LSTM (Long Short Term Memory): Long Short Term Memory networks usually just called "LSTMs" are a special kind of Recurrent Neural Network(RNN)[6], capable of learning long-term dependencies. They were introduced by Hoch Reiter & Schmidhuber (1997). LSTMs are explicitly designed to avoid the long-term dependency problem.

Remembering information for long periods of time is practically their default behavior, not something they struggle to learn! All recurrent neural networks have the form of a chain of repeating modules of neural network. In standard RNNs, the recurrent neural network structure is a little different from the traditional feed forward NN you may be accustomed to seeing. The feed forward network consists of input nodes, hidden units and output node.

B. Data Cleaning and Preprocessing [6]

- Load in and visualize the data: The system using IMDB movies review dataset. If it is stored in machine in a txt file, then it gets load in it.
- Convert to lower case: The reviews which are provided by user in which some will be in capital letter and some in small so to use in a standard way the review will be converted in lowercase.
- Remove punctuation: special character such as @,#,$,etc. are removed from the sentence and will use for preprocessing of data.
- Create list of reviews: Here the system will receive many strings into one string. Now the system separate out individual reviews and store them as individual list elements.
- Like, [review_1, review_2, review_3……. review_n].
- Tokenize (Create Vocab to Int mapping dictionary): In most of the NLP tasks, first create an index mapping dictionary in such a way that which are frequently occurring words are assigned lower indexes.
- Tokenize (Encode the words): It is done by two ways and that are mentioned below
- List of reviews
- Index mapping dictionary using vocab from all reviews. All this above process is used to create an encoding of reviews (replace words in reviews by integers)
- Tokenize (Encode the labels): This is simple because it has two output labels. In which first label is 'positive' and it is assigned by '1' and another label i.e. 'negative' is assigned as '0'.
- Padding / Truncating the remaining data: To deal with both short and long reviews, it gets pad or truncate all reviews to a specific length. It defined the length by Sequence Length. This sequence length is same as number of time steps for LSTM layer.
- Training, Validation, Test Dataset Split: Once data is inserted into a system then it is categorized as training, validation and test sets.
- Dataloaders and Batching: After creating training, test and validation data. Next step is to create dataloaders for this data. For this use generator function for batching data into batches instead it uses a Tensor Dataset. This is one of a very useful utility in PyTorch for using data with Data Loaders with exact same ease as of torch vision datasets.

International Journal of Research in Engineering, Science and Management
Volume-3, Issue-2, February-2020
www.ijresm.com | ISSN (Online): 2581-5792
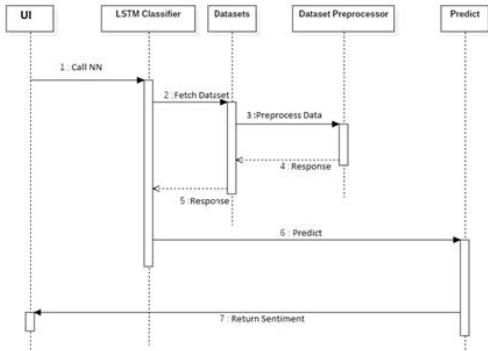
465

C. Design

Sequence Diagram

Fig. 1. Sequence diagram for product rating system.
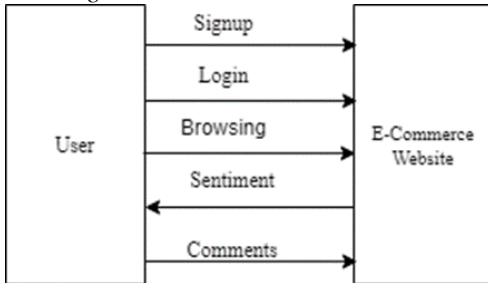
Data Flow Diagram

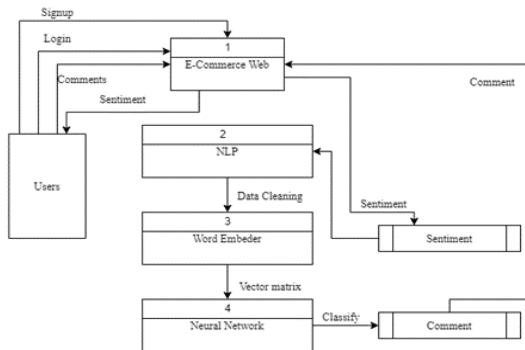Fig. 2. Level 0 data flow diagram (DFD) for product rating system

D. Sequence Diagram

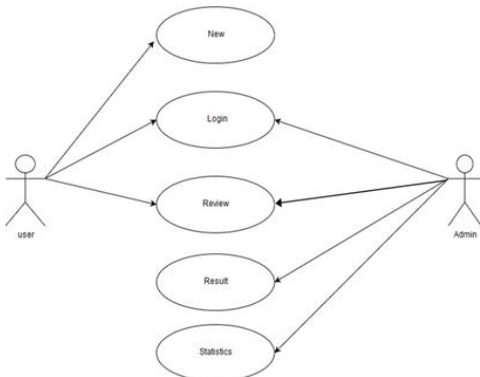Fig. 3. Level 1 data flow diagram (DFD) for product rating system

Fig. 4. Use case diagram for product rating system

4. Results

The output of a proposed system:

5. Future scope

Sentiments which are extracted from the users review it can be easily understand by displaying it in graph form. To extract accurate sentiment LSTM model should be trained with large

International Journal of Research in Engineering, Science and Management
Volume-3, Issue-2, February-2020
www.ijresm.com | ISSN (Online): 2581-5792

466

number of data. It will help company for an effective marketing strategies for predicting about product.

6. Conclusion

The proposed system is designed in such a way that it is able to tackle the problem which are appear in traditional sentimental analysis system and due to this speed, performance and reliability is not get compromised. It provides accurate data and it is less biased because it's accuracy is higher than traditional sentimental analysis system. By using this system companies will be able to know that their product is liked by customers in a market or not and based on this they will take a decision and it will help to do efficiently marketing strategies which will be beneficial. The customer/consumer will get an idea for purchasing a product and it will help to customer to pick a right product and right brand which is more suitable as per their requirement.

References

[1] Raheesa Safrin, K. R. Sharmila, T. S. Shri Subangi, E.A. Vimal "Sentimental Analysis on online product review", 2017.
[2] Sujuta L. Sonavane, "Extracting a sentiment from Review: A Lexicon Based Approach," 2017.
[3] "Comparative Analysis of Sentimental Orientation using SVM and Naïve Bayes Techniques" by Shweta Rana, 2016.
[4] Aarti A. Patil, "Sentimental Analysis of Product review," 2014
[5] The AI University
https://www.youtube.com/channel/UCv6Uw36LRbYnX4HDxKPguKg
[6] https://towardsdatascience.com/sentiment-analysis-using-lstm-step-by-step-50d074f09948.